THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1993 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 268, No. 32, Issue of November 15, pp. 24402-24407, 1993
Printed in U.S.A.

α-Amylase from the Hyperthermophilic Archaebacterium *Pyrococcus furiosus*

CLONING AND SEQUENCING OF THE GENE AND EXPRESSION IN *ESCHERICHIA COLI**

(Received for publication, August 24, 1992, and in revised form, July 28, 1993)

Kenneth A. Laderman‡, K. Asada§, T. Uemori§, H. Mukai§, Y. Taguchi§, I. Kato§, and Christian B. Anfinsen‡¶

From the ‡Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218 and §Takara Shuzo Co., Seta 3-4-1, Otsu, Shiga 520-21, Japan

A gene encoding a highly thermostable α-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus* was cloned and expressed in *Escherichia coli*. The nucleotide sequence of the gene predicts a 649-amino acid protein with a calculated molecular mass of 76.3 kDa, which corresponds well with the value obtained from purified enzyme using denaturing polyacrylamide gel electrophoresis. The NH_2 terminus of the deduced amino acid sequence corresponds precisely to that obtained from the purified enzyme, excluding the NH_2-terminal methionine. The amylase expressed in *E. coli* exhibits temperature-dependent activation characteristic of of the original enzyme from *P. furiosus*, but has a higher apparent molecular weight which is attributed to the improper formation of the native quaternary structure. No homology was found with previously characterized promotor or termination sequences. The deduced amino acid sequence displayed strong homology to the α-amylase A of *Dictyoglomus thermophilum*, an obligately anaerobic, extremely thermophilic bacterium. Evolutionary implications of this homology are discussed.

Hyperthermophilic archaebacteria provide an extraordinary opportunity to study the factors influencing protein thermostability. Unfortunately, due to the recentness of their discovery, the extent of the research completed in this area remains limited. *Pyrococcus furiosus* is an anaerobic marine heterotroph with an optimal growth temperature of 100 °C, isolated by Fiala and Stetter (1986) from solfataric mud off the coast of Vulcano island, Italy. α-Amylase activity has been reported in the cell homogenate and growth medium of *P. furiosus* (Brown *et al.*, 1990; Koch *et al.*, 1990), and the enzyme has been purified to homogeneity (Laderman *et al.*, 1993). The amylase is a homodimer with a molecular mass of 130 kDa which exhibits optimal activity at the optimal growth temperature of the organism. In an attempt to better understand the mechanisms of the enzyme's inherent thermostability the gene coding for the α-amylase from *P. furiosus* was cloned and expressed in *Escherichia coli*, and the nucleotide sequence was determined. In addition the 3′- and 5′-noncoding regions were analyzed in an attempt to identify sequences involved in transcriptional regulation, and a search for homology between the deduced amino acid sequence and other α-amylases was completed.

MATERIALS AND METHODS

Bacterial Strains—Cultures of *P. furiosus* (DSM 3638) were grown as described previously (Ladermanet *al.*, 1993). For cloning and expression of the α-amylase gene, *E. coli* strain JM109 {*rec* A1, Δ(*lac-pro* AB), *end* A1, *gyr* A 96, *thi*-1, *hsd* R 17, *rel* A1, *sup* E 44,/F′*tra* D 36, *pro* AB^+, *lac* $I^qZ\Delta M$ 15} was used.

Plasmids, Enzymes, and Chemicals—The vector pUC18 was used for cloning and DNA sequencing. For expression the vector pTV118N from Takara Shuzo Co. (Kyoto, Japan) was used.

Restriction endonucleases, alkaline phosphatase, DNA Blunting Kit, Random Primer DNA Labeling Kit, DNA Ligation Kit, 7-DEAZA Sequencing Kit, Mutan-K site-directed mutagenesis system, and SeaKem Ultrapure Agarose, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside were obtained from Takara Shuzo Co. (Kyoto, Japan). Geneclean II Kit was obtained from Bio 101 (La Jolla, CA). PCR[1] reagents and enzymes were from Perkin-Elmer Cetus. Hybond-N nylon hybridization filters and [γ-^{32}P]ATP were obtained from Amersham Corp. Ingredients for *E. coli* media were from Difco. Isopropyl-β-D-thiogalactopyranoside (IPTG) and ampicillin were obtained from Sigma. Ultrapure Urea was obtained from Bio-Rad.

Assay of Amylase Activity and Native Polyacrylamide Gel Electrophoresis—The dextrinizing activity of α-amylase was determined at 92 °C using the I_2/KI method as described previously (Laderman *et al.*, 1993). One unit of the enzyme activity was defined as the amount which hydrolyzed 1 mg of starch/min. Native gel electrophoresis and subsequent staining techniques were performed as described elsewhere (Laderman *et al.*, 1993).

Preparation of Chromosomal DNA from P. furiosus—The cells were harvested and approximately 0.1 g of cells (wet weight) was suspended in 0.5 ml of 0.05 M Tris-HCl (pH 8.0) containing 25% sucrose. To the suspension was added 0.1 ml of lysozyme (5 mg/ml). Incubation of the mixture for 1 h at 20 °C was followed by the addition 4 ml of SET solution (150 mM NaCl, 1 mM EDTA, and 20 mM Tris-HCl (pH 8.0)). 0.5 ml of 5% SDS and 100 μl of proteinase K (10 mg/ml) were added, and the mixture was incubated for 1 h at 37 °C. The solution was extracted with phenol-chloroform, and the DNA was precipitated with 2 volumes of ethanol. DNA was recovered by winding and it was rinsed in 80% ethanol. The yield of genomic DNA was approximately 1.3 mg from 0.1 g of cells.

Preparation of a DNA Probe Using PCR—The NH_2-terminal sequence of the intact α-amylase as well as a peptide fragment, determined previously (Laderman *et al.*, 1993), were used for the construction of three degenerate oligonucleotide probes (Fig. 1). The probes were synthesized using an Applied Biosystems model 380B DNA synthesizer.

The PCR reactions were performed using 500 ng of *P. furiosus*

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L22346.

¶ To whom correspondence should be addressed: Dept. of Biology, The Johns Hopkins University, 34th and Charles Sts., Baltimore, MD 21218. Tel.: 410-516-8552; Fax: 410-516-5213.

[1] The abbreviations used are: PCR, polymerase chain reaction; IPTG, isopropyl-β-D-thiogalactopyranoside; kb, kilobase(s).

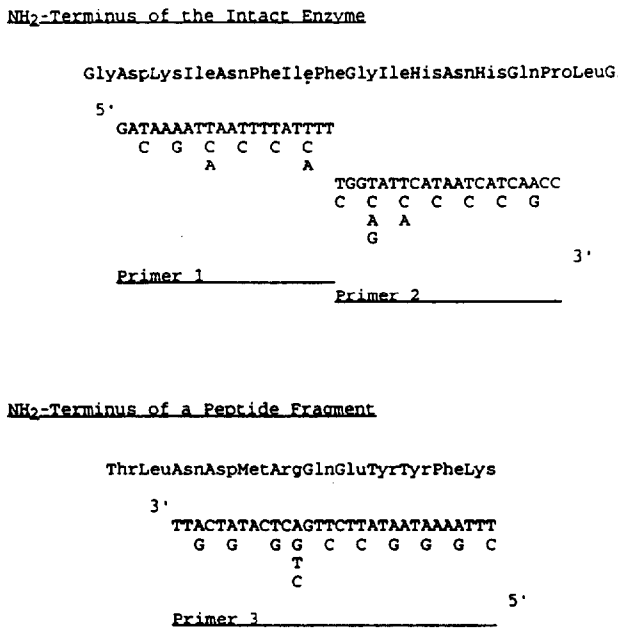

FIG. 1. **Degenerate oligonucleotide primers based on the NH_2-terminal amino acid sequence.** Probes were designed for use in the PCR amplification of a portion of the *P. furiosus* α-amylase gene. The oligonucleotides were prepared as shown, based on the NH_2-terminal sequence of the purified enzyme and a peptide fragment.

chromosomal DNA as the template with 100 pmol of primer 1 and 100 pmol of primer 3 (see Fig. 1). The PCR profile was 94 °C for 0.5 min, 40 °C for 2 min, and 72 °C for 2 min. Amplification was continued for 35 cycles in a total volume of 100 μl. 1 μl of the reaction mixture was further reamplified with primer 2 (see Fig. 1) and primer 3. The PCR profile was same as above, but the number of cycles was 30. Five microliters of this mixture were analyzed by agarose gel electrophoresis.

Cloning and Sequencing of the α-Amylase Gene of P. furiosus—5-μg portions of *P. furiosus* chromosomal DNA were digested with *Pst*I, *Hin*dIII, *Xho*I, or *Eco*RI. The resulting fragments were separated on a 1% agarose gel, transferred to a nylon membrane, and hybridized to the random primer labeled PCR product (Sambrook *et al.*, 1986). Hybridization was carried out for 2 h at 65 °C in 6 × SSC, 0.1% SDS, 5 × Denhardt's solution, 100 μg/ml calf thymus DNA, and 1×10^7 cpm/ml ^{32}P-labeled probe. The membrane was washed for 40 min in a solution of 2 × SSC and 0.1% SDS at 65 °C and then for 20 min in a solution of 0.5 × SSC and 0.1% SDS at 65 °C.

Genomic DNA digested with *Pst*I was size-fractionated on a 1% agarose gel, and a size-fractionated library, with an insert size of approximately 5–5.5 kb, was constructed in the *Pst*I site of pTV118N and used to transform *E. coli* JM109 cells. Recombinant plasmids containing the target sequence were screened by colony hybridization (Sambrook *et al.*, 1986) using the ^{32}P-labeled PCR product produced as described above. The screening yielded three positive clones, and one of these, pKENF, was used for further characterization.

Double-stranded recombinant plasmids and single-stranded DNA were isolated following the protocol of Sambrook *et al.* (1986). Sequencing was completed using the dideoxy termination method of Sanger *et al.* (1977) with the 7-DEAZA sequencing kit. Overlapping subfragments were generated using the method of Yanisch-Perron *et al.* (1985).

Expression of P. furiosus α-Amylase Gene in E. coli—To prepare a construct which expressed the *P. furiosus* α-amylase in *E. coli* an *Nco*I site was created at the translation initiation codon of the α-amylase gene using the site-directed mutagenesis system Mutan-K (Takara Shuzo, Kyoto), converting the original initiation codon from GTG to ATG. This plasmid (pKENF-2N) was digested with *Nco*I and self-ligated to eliminate the 5′-noncoding region and to position the gene at a suitable distance from the vector-derived lac promotor and ribosome binding site. The resulting plasmid (pKENF-N) was digested with *Hin*dIII and self-ligated to delete a 3′-noncoding region; the product was then designated pKENF-NH.

E. coli JM 109 cells carrying this plasmid were designated *E. coli* JM109/pKENF-NH and deposited at the Fermentation Research Institute, Agency of Industrial Science and Technology, Japan, as FERM BP-3782.

E. coli JM 109/pKENF-NH cells were grown for 5 h at 37 °C in 5 ml of L broth containing 100 μg/ml ampicillin. The lac promotor was subsequently induced by the addition of 1 mM IPTG. The cells were allowed to grow for an additional 12 h, with vigorous shaking, then collected by centrifugation (7000 × *g* for 10 min), suspended in 200 μl of 50 mM Tris-HCl, pH 7.0. The mixture was sonicated and centrifuged (30 min at 27,000 × *g*). The supernatant was incubated for 10 min at 99 °C and centrifuged again. This supernatant was used as the crude cell extract.

The α-amylase activity was measured using the standard activity assay described above.

Computer Analysis—The search for existing sequences displaying homology to the *P. furiosus* α-amylase gene was completed through GenBank™ using FASTA searches. Predictions of secondary structure and physical characteristics based on deduced amino acid sequence were completed using PC/GENE (IntelliGenetics Inc., Mountain View, CA) or the Wisconsin Genetics Computer Group (WGCG) sequence analysis software package Version 6.0 (Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, WI).

RESULTS

Cloning and Sequencing of the α-Amylase Gene: Characteristics of Coding and Noncoding Regions—The preparation of a probe using nested PCR resulted in the amplification of a DNA fragment of approximately 1 kilobase. The amplified DNA fragment was blunted using the DNA Blunting Kit (Takara Shuzo, Kyoto) and subcloned into the *Hin*cII site of pUC18 (Sambrook *et al.*, 1986).The cloned plasmid was sequenced by the dideoxy method.

When a Southern blot prepared with digested genomic DNA from *P. furiosus* was probed with the ^{32}P-labeled PCR product, a 5.3-kb *Pst*I fragment, a 3.1-kb *Hin*dIII fragment, a 5.3-kb *Xho*I fragment, and two *Eco*RI fragments of 0.7 and 2.4 kb were found to specifically hybridize to the probe.

Three clones carrying an identical 5.3-kb *Pst*I fragment were identified by colony hybridization of an enriched gene bank of *P. furiosus* genomic DNA using the PCR product known to contain the coding region for the protein's NH_2-terminal sequence. One of these clones, shown to contain the α-amylase coding region in the same orientation as the vector-derived lac promotor, was digested with *Hin*dIII and allowed to self-ligate removing a portion of the 3′-noncoding region. The nucleotide sequence of the resulting 3.1-kb insert was determined in both orientations. The restriction map and sequencing strategy are shown in Fig. 2. The complete nucleotide sequence of the 3.1-kb insert is given in Fig. 3.

The α-amylase gene encompasses 1950 nucleotides, with the initiation codon GTG at position 715 (Fig. 3). There is no strong homology, preceding the coding region, with known archaebacterial, eukaryotic, or eubacterial concensus promotor sequences described previously. Immediately upstream of the coding region is the sequence GGTGGA, similar to the putative ribosome-binding site of the glyceraldehyde-3-phosphate dehydrogenase gene of *Pyrococcus woesei* (GAGGT) (Zwickl *et al.*, 1990). The G + C content of the α-amylase gene is 41.6%, slightly higher than the value reported for the total genome of 38% (Fiala and Stetter, 1986). As has been seen in other sequenced genes from extreme thermophiles, A and T are the preferred bases in the third position of the codons (Zwickl *et al.*, 1990).

The five transcripts of the *Sulfolobus* virus-like particle SSV-1 (Reiter *et al.*, 1988b) and the glyceraldehyde-3-phosphate dehydrogenase gene of *P. woesei* (Zwickl *et al.*, 1990) include the sequence TTTTTT in a pyrimidine-rich region directly downstream of the termination codon. A pyrimidine-rich region exists 34 bases 3′ of the termination codon in the

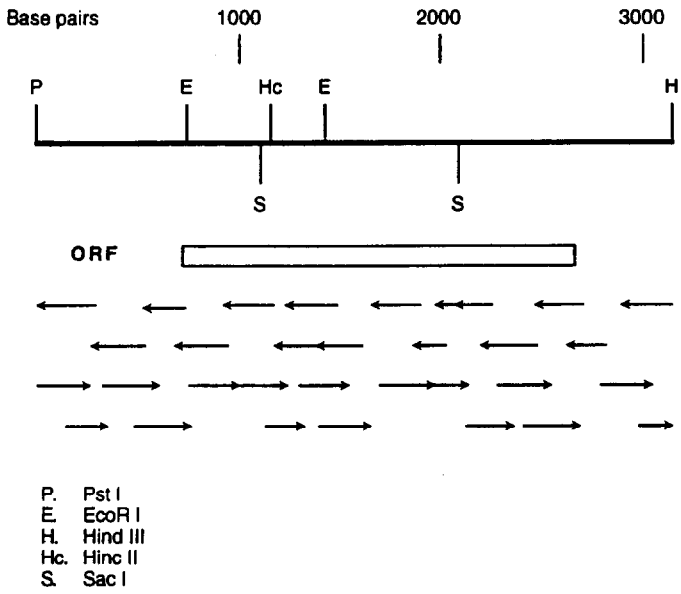

FIG. 2. **Restriction map and sequencing strategy of the cloned *Pst*I-*Hin*dIII insert carrying the α-amylase gene of *P. furiosus*.** *Arrows* indicate the individual sequence runs. *ORF*, open reading frame. *P*, *Pst*I; *E*, *Eco*RI; *H*, *Hin*dIII; *Hc*, *Hinc*II; *S*, *Sac*I.

α-amylase gene, but unlike other archaebacterial sequences examined, there is no pyrimidine-rich region immediately downstream from the TAG stop codon.

Expression of P. furiosus α-Amylase Gene in E. coli and Comparison with the Enzyme Purified from P. furiosus—For expression of the *P. furiosus* α-amylase in *E. coli* an insert containing the gene flanked by archaebacterial noncoding regions was inserted in the expression vector pTV118N. This construct, with the *P. furiosus* 5′ sequence intact, was found to not express any thermophilic α-amylase activity. To prepare a more stream-lined construct for expression in *E. coli*, a unique *Nco*I restriction site was created at the initiation codon, converting the initiation codon from GTG to ATG, and the *P. furiosus* noncoding region was removed. The resultant expression plasmid, denoted pKENF-NH, placed the gene in the correct reading frame at an appropriate distance downstream of the vector promotor (Fig. 4).

IPTG-induced *E. coli* JM109 cells transformed with the plasmid pKENF-NH were found to produce 0.2 unit/ml of thermophilic amylase activity in the crude extract. The heterologously expressed amylase was compared with the enzyme purified from *P. furiosus* regarding molecular weight and temperature dependence of activity. When the apparent molecular weights of the recombinant protein and the isolated enzyme were compared on an activity stained native gel the protein produced in *E. coli* displayed a higher apparent molecular weight. The comparison of the temperature dependence of the α-amylase activity between the purified and recombinant proteins displayed virtually identical relationships of relative activity as a function of temperature (Fig. 5).

When the complete amylase gene was analyzed, no protein or nucleotide sequence was found which displayed complete homology with the deduced NH_2-terminal sequence of the peptide fragment. This suggests that the peptide sequence, on which the synthesis of primer 3 was based, represents a contaminant and not a portion of the *P. furiosus* α-amylase. It was therefore fortuitous that the degenerate primer prepared was able to act as a random primer for the PCR reaction.

To confirm that the α-amylase expressed in *E. coli* was the enzyme purified from the bacterium, not a unique enzyme with a similar activity profile and a shared amino terminal

```
-718                                                    -688
  C TGC AGG GGA GTT TGC AAA GCT AAT AAC TTC AGT AGC TGG TGG AGG AGG CGG AGG AAG
-658                                                    -628
AAA AGA ACT AGC TCA AGG TAA GAT AAG AGA CAT AGA AAA AGC AAA AGA GGC AAT AGA AAA
-598                                                    -568
AGT TAA GGG CTC TCT ATA GCT TTC TAC TCT CCT TCT TTT GGA ATC AGA AAT ATT TCA TAT
-538                                                    -508
TCT GAT CTC CAG AAT GGG AGC TTG TTC ATC TTT ATT TTT ATA TAA TAC TCG GTG CTT TTC
-478                                                    -448
TCT CTG TAT ATT TTC TCC ACT ACT TCC TGG GGC AGT TGG AAC TGA ACT ATA ATT TCA GCA
-418                                                    -388
TCC TCT GAT ACC TTT GTT TCA AAT TTT GAA GTA CCC TTG TAA TCC TTT CCA TTT ACC TTT
-358                                                    -328
ATG AGA TAA TTT GTG CCC TCT GGA AAT TCT ATT TCG AAG TTG AAA TCT GAG ACA ATT TTT
-298                                                    -268
TCC ACT TTT AGC GTA AAT AAC CCC CAA CCG TCT TTT TCA ACT ATT GTA ACT GTC CTG TTT
-238                                                    -208
TCC TTA TAG AAT ATC TCA CTT GAT TCT TTT TCA TTA ATG GTT GCA GGA GGC ATT TTT GCC
-178                                                    -148
GTT CTC ATG GCA AGC ACT AGA AGG ACT ATA AAA ATT ATT GCT ACA GCT GTT ATT TTC TTG
-118                                                    -88
TCC ATG CTA ACA CCC TGT AAT GAG ATT TGG ATT TTC CTA TAT AAA AAG CCT TAG TTA TTT
-58                                                     -28                    -1 1
TTG AGC CAT TAA ATA TAT AAG GAA GTA TCA CTC TTA GTG ATT AAT GGG TGG ACG GAA GTG
                                                                               met
3                                                       33
GGA GAT AAA ATT AAC TTC ATA TTT GGA ATT CAC AAC CAT CAG CCC CTG GGC AAC TTT GGA
gly asp lys ile asn phe ile phe gly ile his asn his gln pro leu gly asn phe gly
63                                                      93
TGG GTG TTT GAG GAG GCT TAT GAA AAG TGT TAC TGG CCG TTT CTG GAG ACT CTG GAG GAA
trp val phe glu glu ala tyr glu lys cys tyr trp pro phe leu glu thr leu glu glu
123                                                     153
TAT CCA AAC ATG AAG GTT GCC ATT CAT ACA AGT GGC CCC CTC ATT GAG TGG CTC CAA GAT
tyr pro asn met lys val ala ile his thr ser gly pro leu ile glu trp leu gln asp
183                                                     213
AAT AGA CCC GAA TAC ATA GAC TTG CTT AGA AGT CTA GTG AAA AGA GGA CAG GTG GAG ATA
asn arg pro glu tyr ile asp leu leu arg ser leu val lys arg gly gln val glu ile
243                                                     273
GTC GTT GCT GGG TTC TAC GAG CCT GTG CTA GCA TCA ATC CCA AAG GAA GAT AGA ATA GAG
val val ala gly phe tyr glu pro val leu ala ser ile pro lys glu asp arg ile glu
303                                                     333
CAG ATA AGG TTA ATG AAA GAG TGG GCT AAG AGT ATT GGA TTT GAT GCT AGG GGA GTT TGG
gln ile arg leu met lys glu trp ala lys ser ile gly phe asp ala arg gly val trp
363                                                     393
CTA ACT GAA AGA GTA TGG CAA CCA GAG CTC GTA AAG ACC CTT AAG GAG AGC GGA ATA GAT
leu thr glu arg val trp gln pro glu leu val lys thr leu lys glu ser gly ile asp
423                                                     453
TAT GTA ATA GTT GAC GAT TAC CAC TTC ATG AGT GCG GGA TTA AGT AAA GAG GAG CTG TAC
tyr val ile val asp asp tyr his phe met ser ala gly leu ser lys glu glu leu tyr
483                                                     513
TGG CCA TAT TAT ACG GAA GAT GGT GGG GAA GTT ATA GCT GTT TTC CCG ATA GAT GAG AAG
trp pro tyr tyr thr glu asp gly gly glu val ile ala val phe pro ile asp glu lys
543                                                     573
TTG AGA TAT TTG ATT CCC TTT AGA CCC GTT GAT AAG GTC TTA GAA TAC CTG CAT TCT CTC
leu arg tyr leu ile pro phe arg pro val asp lys val leu glu tyr leu his ser leu
603                                                     633
ATA GAT GGT GAT GAG AGC AAA GTT GCA GTA TTT CAT GAC GAT GGT GAG AAG TTT GGA ATC
ile asp gly asp glu ser lys val ala val phe his asp asp gly glu lys phe gly ile
663                                                     693
TGG CCT GGA ACT TAT GAG TGG GTG TAT GAA AAG GGA TGG TTA AGA GAA TTC TTT GAT AGA
trp pro gly thr tyr glu trp val tyr glu lys gly trp leu arg glu phe phe asp arg
723                                                     753
ATT TCA AGT GAT GAA AAG ATA AAC TTA ATG CTT TAC ACT GAA TAC TTA GAA AAA TAT AAG
ile ser ser asp glu lys ile asn leu met leu tyr thr glu tyr leu glu lys tyr lys
783                                                     813
CCT AGA GGT CTT GTT TAT CTT CCA ATA GCT TCA TAT TTT GAG ATG AGC GAA TGG TCA TTG
pro arg gly leu val tyr leu pro ile ala ser tyr phe glu met ser glu trp ser leu
843                                                     873
CCA GCA AAG CAG GCA AGG CTC TTT GTG GAG TTC GTC AAT GAG CTT AAA GTT AAA GGT ATA
pro ala lys gln ala arg leu phe val glu phe val asn glu leu lys val lys gly ile
903                                                     933
TTT GAA AAG TAC AGG GTA TTT GTT AGG GGA GGA ATT TGG AAG AAT TTC TTC TAT AAA TAC
phe glu lys tyr arg val phe val arg gly gly ile trp lys asn phe phe tyr lys tyr
963                                                     993
CCA GAG AGC AAC TAC ATG CAC AAG AGA ATG CTA ATG GTA AGT AAG TTA GTG AGA AAC AAT
pro glu ser asn tyr met his lys arg met leu met val ser lys leu val arg asn asn
1023                                                    1053
CCT GAG GCC AGG AAG TAT CTG CTG AGA GCA CAA TGT AAC GAT GCT TAT TGG CAC GGC CTC
pro glu ala arg lys tyr leu leu arg ala gln cys asn asp ala tyr trp his gly leu
1083                                                    1113
TTC GGT GGA GTA TAT TTA CCC CAT CTT AGG AGG GCC ATC TGG AAC AAT TTA ATC AAG GCC
phe gly gly val tyr leu pro his leu arg arg ala ile trp asn asn leu ile lys ala
1143                                                    1173
AAC AGC TAT GTA AGC CTT GGA AAG GTC ATA AGG GAT ATC GAC TAC GAT GGC TTT GAG GAA
asn ser tyr val ser leu gly lys val ile arg asp ile asp tyr asp gly phe glu glu
1203                                                    1233
GTT CTC ATA GAG AAT GAC AAC TTT TAT GCA GTG TTT AAA CCC TCT TAC GGT GGT TCC TTG
val leu ile glu asn asp asn phe tyr ala val phe lys pro ser tyr gly gly ser leu
1263                                                    1293
GTG GAG TTT TCA TCA AAG AAT AGA CTC GTG AAT TAT GTA GAT GTT CTG GCA AGA AGG TGG
val glu phe ser ser lys asn arg leu val asn tyr val asp val leu ala arg arg trp
1323                                                    1353
GAA CAC TAT CAT GGC TAT GTG GAA AGT CAA TTT GAT GGA GTA GCC AGC ATT CAT GAG CTC
glu his tyr his gly tyr val glu ser gln phe asp gly val ala ser ile his glu leu
1383                                                    1413
GAG AAA AAG ATA CCA GAT GAA ATA AGA AAA GAA GTT GCT TAC GAC AAG TAC AGA AGG TTC
glu lys lys ile pro asp glu ile arg lys glu val ala tyr asp lys tyr arg arg phe
1443                                                    1473
ATG CTT CAA GAT CAC GTA GTC CCC CTG GGA ACA ACT CTG GAA GAC TTC ATG TTC TCA AGA
met leu gln asp his val val pro leu gly thr thr leu glu asp phe met phe ser arg
1503                                                    1533
```

FIG. 3. **Nucleotide and deduced amino acid sequence of the α-amylase gene.** The nucleotide sequence from the *Pst*I site, 717-bp 5′ of the initiation codon, to the *Hin*dIII site at position 2423 is presented. The *underlined* portion represents the nucleotide sequence coding for the NH_2-terminal amino acids upon which the oligonucleotide primers were based. The *single underlined* region represents the sequence homologous to primer 1. The *double underlined* region represents the sequence homologous to primer 2.

```
CAA CAG GAG ATC GGA GAG TTT CCT AGG GTT CCA TAC TCA TAT GAA CTA CTA GAT GGA GGA
gln gln glu ile gly glu phe pro arg val pro tyr ser tyr glu leu leu asp gly gly
1563                                    1593
ATA AGG CTG AAG AGG GAA CAC TTG GGA ATA GAA GTT GAA AAA ACA GTG AAG TTA GTG AAT

ile arg leu lys arg glu his leu gly ile glu val glu lys thr val lys leu val asn
1623                                    1653
GAT GGA TTT GAG GTG GAG TAT ATA GTG AAC AAC AAG ACA GGA AAT CCT GTA TTG TTC GCA
asp gly phe glu val glu tyr ile val asn asn lys thr gly asn pro val leu phe ala
1683                                    1713
GTG GAA CTT AAC GTT GCA GTT CAG AGC ATA ATG GAG AGC CCA GGA GTT CTA AGG GGG AAA
val glu leu asn val ala val gln ser ile met glu ser pro gly val leu arg gly lys
1743                                    1773
GAA ATT GTC GTT GAT GAC AAG TAT GCA GTT GGG AAG TTT GCA CTG AAG TTT GAA GAC GAA
glu ile val val asp asp lys tyr ala val gly lys phe ala leu lys phe glu asp glu
1803                                    1833
ATG GAA GTC TGG AAG TAT CCA GTA AAG ACT CTC AGT CAA AGT GAA AGT GGC TGG GAT CTA
met glu val trp lys tyr pro val lys thr leu ser gln ser glu ser gly trp asp leu
1863                                    1893
ATC CAG CAG GGT GTC AGC TAC ATA GTT CCA ATA AGG TTG GAG GAT AAA ATA AGG TTT AAG
ile gln gln gly val ser tyr ile val pro ile arg leu glu asp lys ile arg phe lys
1923                                    1953
CTA AAA TTT GAG GAA GCC TCG GGA TAG GGA GGC CCT CAT CAC CAA TCA GGG CCC GAA AGA
leu lys phe glu glu ala ser gly AMB
1983                                    2013
CTC CCT CAT CGG CCC TTC TAT TTT ATT TTA AAC GTC AAT GGT TTA CCA AGT TTC CAA AAC
2043                                    2073
TTA CAA AAT GAA CAA ATC TCT CCA CTT GCG GGC ATT CCA CAT ATC TTG CAC TCT TTG AGG
2103                                    2133
TCT TTC CCC TTC ACT TCT GGC TCG AAA AGT TTT TTC TTT CTT AGG AAT CCT CTC ACG AAG
2163                                    2193
TTG AAC TTT GTT CCA GGC CTT TTT TCC TCC AAT TCA TTG AGA ACT TCC TTC ATG TCA AGA
2223                                    2253
GTT GTC GCA CCT CTT GCA TAA GGA CAC TCC TCT ACT ATG TAC TCC AAT CCA ACG GCA ATG
2283                                    2313
GCA TAG GCA ACA ACT TCC CTC TCA GTT AAT TCG TAG AGA GGT TTG ATC TTC TTT ACG AAC
2343                                    2373
TTT CCT TCC CCT GGG AGC AGA GGA CCT CCC TTA GCC AGG TAC TCT GTA TTC CAG TGG AGT
2403
AAG TTG TTC ATG AGA AAG CTT
```

FIG. 3—*Continued*

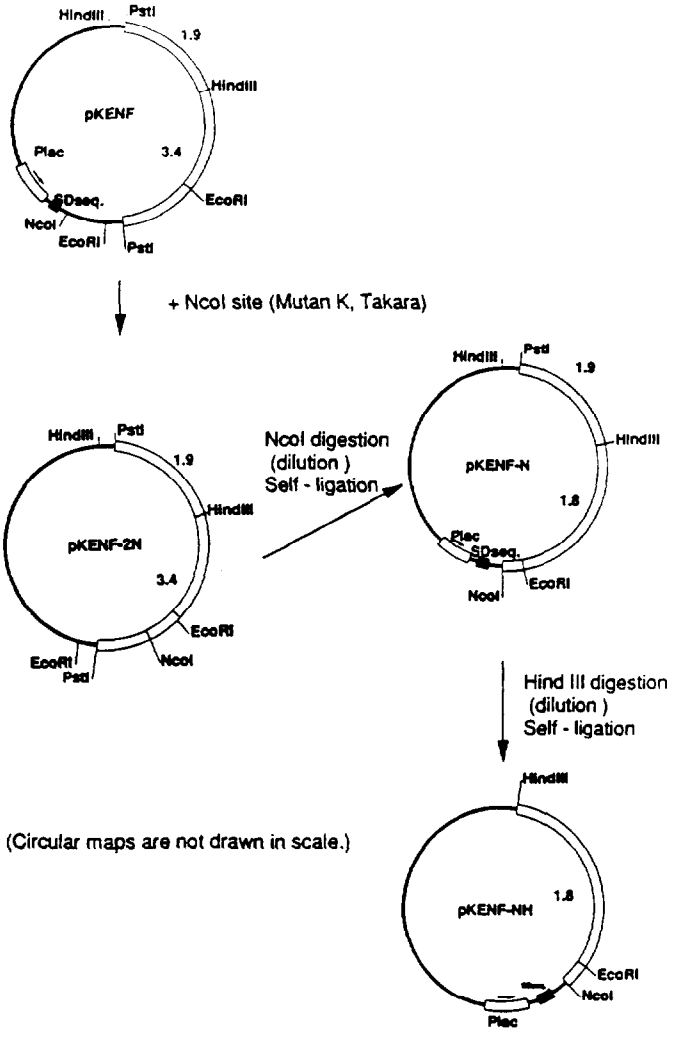

FIG. 4. **The construction scheme for the recombinant plasmid pKENF-NH, used for the expression of the *P. furiosus* α-amylase gene in *E. coli*.** The expression plasmid was constructed by the introduction of a *Nco*I site at the initiation codon of the amylase gene, followed by restriction digestion with this enzyme and subsequent self ligation. The resulting construct positions the initiation codon adjacent to the Shine-Dalgarno sequence of the plasmid lac promotor. The construct was completed by digestion with *Hin*dIII followed by self-ligation, to remove a portion of the 3′-noncoding region.

sequence, the physical characteristics of the protein were examined. The gene product and the *P. furiosus* α-amylase were found to have comparable isoelectric points, specific activities, and pH of optimal activity (data not shown).

Primary Structure of P. furiosus α-Amylase and Computer Aided Comparison with Enzyme Homologs from Mesophilic and Thermophilic Sources—The deduced amino acid sequence of the *P. furiosus* α-amylase comprises 649 amino acids, with a calculated molecular mass of 76.3 kDa. This agrees well with the apparent molecular mass of the protein, determined by gel electrophoresis under denaturing conditions, of 66 kDa (Laderman *et al.*, 1993).

It is known that the amylase of *P. furiosus* is secreted into the growth medium under native conditions (Koch *et al.*, 1990). When the primary sequence of the protein was analyzed using the PC/GENE PSIGNAL program, no typical eubacterial or eukaryotic NH_2-terminal signal sequences were found. Using the standard activity assay, it was not possible to detect any thermophilic amylase activity in the extracellular media of JM109/pKENF-NH cell, confirming the absence of a signal sequence which would make the protein competent for export in *E. coli*. When the NH_2-terminal sequence of the native enzyme purified from *P. furiosus* was compared with the predicted NH_2-terminal sequence they were found to be identical, suggesting there is no NH_2-terminal processing.

The GenBank™ FASTA searches resulted in the acquisition of only a single strongly homologous protein sequence. The query with the protein sequence of the *P. furiosus* α-amylase, using the method of Pearson and Lipman (1988), identified a 41.9% identity in a 537-amino acid overlap in the sequence of one of the α-amylases from the extremely thermophilic bacterium *Dictyoglomus thermophilum* designated α-amylase A (Fukusumi *et al.*, 1988). The complete sequence of the *D. thermophilum* α-amylase and a number of additional α-amylases from various sources were aligned using the PC/GENE CLUSTAL multiple sequence alignment program, and none displayed the high homology noted above. When a search for homology with previously identified consensus sequences, known to be located in the active center and participate in substrate binding in a number of amylases from a variety of sources (Bahl *et al.* 1991; Tsukagoshi *et al.*, 1985), was performed no significant homology was found.

The codon usage of α-amylases from three thermophilic sources, *P. furiosus*, *D. thermophilum* (Fukusumi *et al.*, 1988), and *Bacillus stearothermophilus* (Tsukagoshi *et al.*, 1985), were compared. As noted previously, the higher the optimal temperature of activity the more extensive the bias against the usage of the dinucleotide CG (Zwickl *et al.*, 1990). This trend is apparent not only for the arginine codons, but for the serine, proline, threonine, and alanine codons as well. No other anomalies in codon usage attributable to thermostability are apparent other than shifts inherent to the changes in amino acid composition.

The *D. thermophilum* α-amylase A displays physical characteristics similar to those observed with the *P. furiosus* α-amylase. The *D. thermophilum* enzyme exhibits optimal activity at 90 °C with approximately 70% residual activity following incubation for 1 h at this temperature (Fukusumi *et*

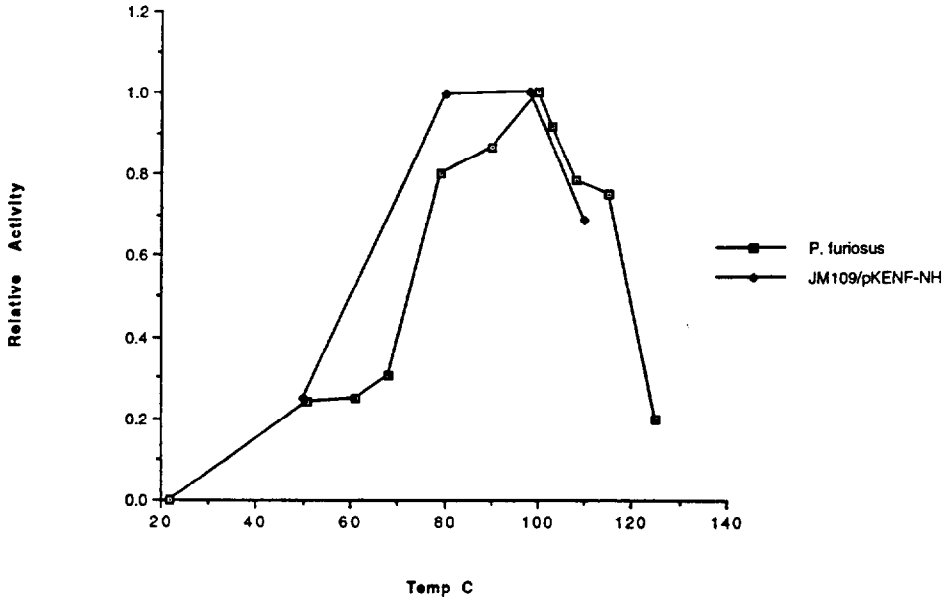

FIG. 5. **Comparison of the temperature-dependent amylase activity of the *P. furiosus* α-amylase and the amylase activity of the heat-treated crude extract from pKENF-NH-transformed JM109 *E. coli* cells.** Amylase activity was determined at a variety of temperatures, using the standard technique.

al., 1988). In contrast, the *P. furiosus* amylase is optimally active at 100 °C and exhibits a substantially higher thermostability: 85% after 3 h at 100 °C (Laderman *et al.*, 1992). The variance in temperature-dependent activity and thermostability between two proteins displaying a high level of homology provides a unique opportunity to investigate the aspects of the primary sequence which confer enzyme thermostability.

Computer analysis of the primary structures and the predicted secondary structures were prepared for the α-amylases from *P. furiosus* and *D. thermophilum* in an attempt to identify possible factors effecting thermostability. When hydropathy of the two sequences was plotted, using the PC/GENE SOAP program, no apparent increase was found in the overall hydrophobicity associated with the increase in the thermostability of the *P. furiosus* amylase. When regions of high primary sequence homology were compared, no specific trend in hydropathy was noted.

When the predicted secondary structures of the two proteins (obtained using the PC/GENE GARNIER program) were compared, little similarity in the proposed structures was noted. This dissimilarity was seen both overall and in areas of high primary structure homology. It is not possible to deduce, with confidence, the protein's secondary structure based exclusively on computer analysis. These results indicate that the primary sequence motifs upon which the computer secondary structure predictions are based differ sufficiently between the two proteins.

DISCUSSION

A number of unusual characteristics of the *P. furiosus* α-amylase gene set it apart from a majority of the genes characterized to date. The gene utilizes the relatively rare initiation codon GTG, as does the glyceraldehyde-3-phosphate dehydrogenase gene from *P. woesei* (Zwickl *et al.*, 1990). It is possible that this represents a tendency in the usage of this initiation codon in hyperthermophilic archaebacteria. Unfortunately the number of structural genes isolated from these sources are too limited to allow an accurate assessment of whether the initiation codon GTG is in fact a preferred initiation codon in these organisms, although this may be an intriguing possibility.

It is known that the production of α-amylase from *P. furiosus* is increased by the presence of starch (data not shown) indicating that the gene possesses an inducible promotor, a feature previously uncharacterized in hyperthermophilic archaebacteria. When the 5′-noncoding region of the amylase gene was compared with the promotor sequences of other archaebacteria, no homology with the consensus sequences previously identified in *Sulfolobus* and *Methanococcus* (Reiter *et al.*, 1988a) was found. The lack of homology with the putative ribosome-binding site of the *P. woesei* glyceraldehyde-3-phosphate dehydrogenase gene suggests either a difference in the 3′ terminus of the 16 S rRNA between the two closely related species, or a different promotor mechanism or both. In addition the *P. furiosus* amylase gene lacks the pyrimidine-rich region found immediatly downstream of the proteins 3′ termini in the forementioned archaebacterial genes. This lack of homology with previously investigated archaebacterial promoters and termination sequences may be a result of the local environment of the gene. The *P. woesei* glyceraldehyde-3-phosphate dehydrogenase gene is flanked closely by a number of open reading frames, which would benefit from a translational coupling mechanism. In two instances with the SSV1 genes in *Sulfolobus*, shown to share similar termination sequence characteristics with the *P. woesei* gene, this linkage between termination and re-initiation was observed. The *P. furiosus* gene exhibits no flanking open reading frames within the insert which was sequences, therefore precluding the necessity for translational coupling. This characteristic as well as the inducibility of the gene are unique among the archaebacterial genes investigated thus far.

The α-amylase from *P. furiosus* is one of a number of thermophilic proteins which have been expressed, in mesophilic hosts, in an active form. The three forms of α-amylase from *D. thermophilum* (Fukusumi *et al.*, 1988), xylan-degrading enzymes from *Caldocellum saccharolyticum* (Luthi *et al.*, 1990), and the glyceraldehyde-3-phosphate dehydrogenase from *P. woesei* (Zwickl *et al.*, 1990), produced under *in vivo* conditions at 73, 70, and 100 °C, respectively, have all been successfully expressed in *E. coli*. In these instances, the proteins produced by the transformed bacteria remained inactive until they were heated to the temperature appropriate to the enzyme's native conditions. This suggests that the proper folding of the protein into a structure which can be temperature-activated is possible at temperatures other than those at which the protein is produced under native growth conditions.

The *P. furiosus* α-amylase which is a dimer differs from the aforementioned examples in that it additionally requires the formation of an appropriate quaternary structure. Although temperature-dependent amylase activity was observed in *E. coli*, the apparent native molecular weight of the enzyme was higher than the form purified from *P. furiosus*, suggesting improper subunit assembly. It is not possible, however, to determine whether this improper assembly is due to translation at lower temperature or to unidentified aspects of production in *E. coli*.

Perhaps the most interesting facet of this research are the evolutionary implications of the sequence homology between the α-amylases of *P. furiosus* and *D. thermophilum*. Based on molecular phylogeny using rRNA sequences, existing organisms are seen to fall into three coherent groups eukaryotes, eubacteria, and archaebacteria (Fox *et al.*, 1980). Substancial physiological and structural differences exist between archaebacteria and eubacteria, which is evidence of their deep evolutionary separation (Woese, 1985). The phylogenetic tree prepared by Pace *et al.* (1986) places archaebacteria closer to the common ancestor of all the kingdoms than one or both of the other primary kingdoms, suggesting that archaebacteria are more primitive than one or both of the other lines. *D. thermophilum* is a Gram-negative, obligately anaerobic, extremely thermophilic bacterium (Saiki *et al.*, 1985). It shares with *P. furiosus* a low G + C content and a tolerance for extreme thermal conditions, but is a member of a different phylogenetic kingdom. *D. thermophilum* has been shown to produce three different species of α-amylase, which can be classified into two separate classes. First, amylase A, which displays a high degree of homology with the *P. furiosus* α-amylase, and second, amylase B and amylase C, which display homology with Taka-amylase A (Toda *et al.*, 1982). No significant homology exists between these two classes, suggesting that they represent two independent gene families or a single family which diverged at a point so distant that no feature save enzyme activity remains as evidence of their relationship. It is possible, since archaebacteria are considered to represent a primitive kingdom, that the *P. furiosus* α-amylase, and therefore the *D. thermophilum* amylase A, may be an example of an archaic form of the enzyme which is well suited to extreme temperatures. In contrast, the *D. thermophilum* amylases B and C contain regions known to be well conserved in several *Bacillus* species, hog, mouse, and human amylases (Fukusumi *et al.*, 1988), and they represent the common form of the enzyme, various examples of which are active over a wide range of temperatures.

REFERENCES

Bahl, H., Burchhardt, G., Spreinat, A., Haeckel, K., Weinecke, A., Schmidt, B., and Antranikian, G. (1991) *Appl. Environ. Microbiol.* **57,** 1554–1559

Brown, S. H., Constantino, H. R., and Kelley, R. M. (1990) *Appl. Environ. Microbiol.* **56,** 1985–1991

Fiala, G., and Stetter, K. O. (1986) *Arch. Microbiol.* **145,** 56–61

Fox, G. E., Staekebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magnum, L. J., Zableno, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Laehrsen, K. R., Chen, K. N., and Woese, C. R. (1980) *Science* **209,** 457–463

Fukusumi, S., Kamizono, A., Horinouchi, S., and Beppu, T. (1988) *Eur. J. Biochem.* **174,** 15–21

Koch, R., Zablowski, P., Spreinat, A., and Antranikian, G. (1990) *FEMS Microbiol. Lett.* **71,** 21–26

Laderman, K. A., Davis, B. R., Krutzsch, H. C., Lewis, M. C., and Anfinsen, C. B. (1993) *J. Biol. Chem.* **268,** 24394–24401

Luthi, E., Love, D. R., McAnulty, J., Wallace, C., Caughey, P. A., Saul, D., and Bergquist, P. L. (1990) *Appl. Environ. Microbiol.* **56,** 1017–1024

Pace, N. R., Olsen, G. J., and Woese, C. R. (1986) *Cell* **45,** 325–326

Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2444–2448

Reiter, W., Palm, P., and Zillig, W. (1988a) *Nucleic Acids Res* **16,** 1–19

Reiter, W., Palm, P., and Zillig, W. (1988b) *Nucleic Acids Res.* **16,** 2445–2459

Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985) *Int. J. Syst. Bacteriol.* **35,** 253–259

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1986) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467

Toda, H., Kondo, K., and Narita, K. (1982) *Proc. Jpn. Acad.* **58,** 208–212

Tsukagoshi, N., Iritani, S., Sasaki, T., Takemura, T., Ihara, H., Idota, Y., Yamagata, H., and Udaka, S. (1985) *J. Bacteriol.* **164,** 1182–1187

Woese, C. R. (1985) *The Bacteria*, Academic Press, New York

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene (Amst.)* **33,** 103–119

Zwickl, P., Fabry, S., Bogedain, C., Haas, A., and Hensel, R. (1990) *J. Bacteriol.* **172,** 4329–4338